

Interactive Analysis of Cancer Registry Data on the World Wide Web

Eric B. Durbin, M.S., Mark Schneider, B.S., Sudhir Nelvagal, M.S.
Kentucky Cancer Registry, University of Kentucky, Lexington, Kentucky

Population based central cancer registries are an integral part of the cancer prevention and control process.¹ Analysis of registry data is used to identify high incidence rates, plan interventions, target resources and measure efficacy over time. A statewide registry database has been maintained by the Kentucky Cancer Registry since 1991. The database application was implemented using Raima Data Manager. A complete set of analytic tools were developed for use with this database. These tools have been modified to provide interactive data analysis of the registry database in real time using a World Wide Web browser. By making this tool available on the Web the use of central registry data will be highly accessible to a larger community involved in cancer control.

The Web browser functions as the front end to the cancer registry database. The original database structure has not been modified. All confidential patient identifiers have been removed. An interactive data analysis session involves five steps: login to the database, build an SQL query to select a population (i.e. all females with late stage breast cancer), submit the query to the database, apply analytic functions to the queried records to generate reports, tables, and graphs, and logout to end the session. SQL queries can also be saved and retrieved for later use.

Four major components were developed for Web based interactive data analysis. A Common Gateway Interface (CGI) engine provides a menu driven interface for the Web browser, a name server coordinates communications between the CGI engine and the database, a database server manages all access to the database, and a Hypertext Markup Language (HTML) engine formats database output for the Web browser. Interprocess communications between components are provided through the use of TCP/IP sockets.

The CGI engine consists of C programs which are executed on the Web server. A menu driven interface emulates the original database application. Dynamic menus are generated "on the fly" depending upon options selected by the user. The CGI engine generates HTML forms for data input. It is also used to build the SQL queries and presents database results

to the browser. The CGI engine submits login, logout, analytic requests, and SQL queries directly to the database server. From the database server it receives HTML formatted output.

The name server also resides on the Web server. It coordinates communications between the CGI and database engines. The name server maintains the information necessary for establishing secured connection oriented sessions between the CGI engine and the database server.

The database server accesses the cancer registry database. It authenticates a user login and establishes a child process to handle all subsequent database requests. The child process is assigned a unique identifier known only to the name server. All database requests from the CGI engine can therefore be secured through the name server. The database server executes queries, performs analytic calculations, and stores and retrieves queries. The child process exits upon logout.

The HTML engine is invoked by the database server to format graphical and text based data for submission to the Web browser. All output is formatted in the HTML language. The HTML engine formats tables, text files, and creates graphs in the form of GIFs.

A prototype tool was successfully implemented to perform real time interactive data analysis of cancer registry data using a Web browser. Connection oriented sessions were achieved using information maintained in the name server. Interactive construction of highly complex SQL queries was hindered by limitations in the HTML language. Therefore, new technologies such as Java are currently being explored.

Hardware: Sun SPARCServer 20, Operating System: Solaris 2.5, Database Engine: Raima Data Manager 3.21a, Web Server: NCSA 1.5, Web Browser: Netscape 2.1. URL: <http://www.kcr.uky.edu>.

1. Friedell, G.H., Tucker, T.C., Prevention and Control, "Central Cancer Registries: Design, Management, and Use" Ed. Herman Menck and Charles Smart. 1994.